数理解析研究所講究録
494巻 1983年 35-45

分散形データベースにおける単純問合せの最適化について

Simple Query Processing in Star Computer Networks

広島大学　工学部　杉原　一夫　(Kazuo SUGIHARA)
宮尾　淳一　(Jun'ichi MIYAO)
菊野　亨　(Tohru KIKUNO)

1. まえがき

分散形データベースシステムにおいて，問合せ処理のための最適なスケジュールを求めることは基本的かつ重要な問題である．一般の問合せに対し最適なスケジュールを求める問題はＮＰ困難であることが既に示されている[7]．そこで，最適なスケジュールを効率良く求めることができる問合せのクラスに興味が集まってきている．その1つに単純問合せ（simple query）というクラスがある[2],[5],[7]．例えば，ネットワークの形状を完全グラフと見なすときに各サイト間の通信コストがすべて等しい場合，単純問合せに対する通信コスト最小のスケジュールは $O(n \log n)$ の手数で求まることが示されている[2]．ここで，n はサイトの総数である．

本稿では，星状ネットワークにおける単純問合せに対し，応答時間を最小にする問題ＲＴと通信コスト及び計算コスト

の総和を最小にする問題TCについて考察し，それぞれ $O(n^2)$，$O(n \log n)$ の手数で解けることを示す．

2. 準備

2.1 システムモデル

分散形データベースシステムはコンピュータネットワークとその上に分散配置されたデータベースから構成されている．本稿では，コンピュータネットワークの形状は星状であり，且つデータベースは関係データベースであるとする．

コンピュータネットワークを無向グラフ $G=(V, E)$ で表す．ここで，各節点 $v_i \in V$ はネットワークのサイトに，各無向枝 $(v_i, v_j) \in E$ はサイト v_i，v_j 間の通信回線にそれぞれ対応する．サイト v_0 をセンタ（center）とする星状ネットワーク（star network）$G=(V, E)$ とは，

$$V=\{v_0, v_1, v_2, \cdots, v_n\}$$

$$E=\{(v_0, v_i) \mid 1 \leq i \leq n\}$$

を満たすネットワークである．ここで，n は正整数で，G におけるセンタ以外のサイトの数を表している．

関係データベース[6]とは関係（relation）の集合である．通常，関係は2次元配列の形をした表として表現される．表の各列は異なる属性でラベル付けされ，各列の値はその列に

対応する属性の値域(domain)から選ばれる．一方，表の各行はその表に対応する属性の集合から各値域への1つの写像である．表の1つの行を組(tuple)と呼ぶ．以下，関係r_iの組の数をr_iのサイズと呼び，$|r_i|$で表す．

2.2 問合せ

分散形データベースシステムにおける問合せ処理は，一般に，各サイトが他のサイトからのデータ転送を必要とせずに独自に実行できる部分と，他のサイトからのデータ転送を終えた後に初めて実行できる部分とに分けられる．前者をローカル処理，後者をグローバル処理と呼ぶ．例えば，関係演算の中で，選択と射影はローカル処理と見なせるが，自然結合は一般にグローバル処理となる．

一般的な問合せの部分クラスとして，ある属性Aに対する単純問合せ(simple query)を次のように定義する．

[定義1] 単純問合せとは関係$r = r_1 \bowtie r_2 \bowtie \cdots \bowtie r_m$を求める問合せである．ここで，$\bowtie$は自然結合[6]を表し，各$r_i$ $(1 \leq i \leq m)$はサイトv_iにおいてローカル処理後に得られた関係で，且つ，共通の属性A上の関係である．なお，rはDサイトと呼ばれる特定のサイト$v_d \in V$に転送されるものとする．

本稿では，次の仮定の下に議論する．単純問合せ$r = r_1 \bowtie$

$r_2 \bowtie \cdots \bowtie r_m$ における関係 r_i, r_j ($1 \leq i, j \leq m$) に対し，$|r_i \bowtie r_j| = |r_i| \times |r_j| \times \frac{1}{D}$ とする．ここで，Dは単純問合せにおける共通の属性Aに対応する値域の要素数とする．

単純問合せ処理の最適化は，定義1より，ローカル処理の最適化とグローバル処理の最適化に分けることができる．ローカル処理の最適化に対しては，従来の集中形データベースシステムにおける問合せ処理の最適化の技法[(6)]が適用できる．従って，本稿ではグローバル処理の最適化についてのみ議論する．

2.3 スケジュールグラフ

単純問合せに対するグローバル処理の手順を記述するために，次のように定義されるスケジュールグラフを導入する．

[定義2]　ネットワーク $G=(V, E)$ 上でのスケジュールグラフとは次の条件(1)－(3)を満たす連結な非サイクリック有向グラフ $\mathcal{S}=(\mathcal{N}, \mathcal{A})$ である．

(1) 各節点 $u \in \mathcal{N}$ にはG上のサイト名 $v_i \in V$ がラベルとして付けられている．このラベルを $\ell(u)$ で表す．

(2) 各有向枝 $(u, w) \in \mathcal{A}$ に対し，$\ell(u) \neq \ell(w)$ ならばG上に無向枝 $(\ell(u), \ell(w)) \in E$ が存在する．

(3) そこから出る有向枝が存在せず且つ $\ell(u_d) = v_d$ な

る節点 $u_d \in N$ が $\mathcal{S}$ に唯一つ存在する．u_d をD節点と呼ぶ．

次に，スケジュールグラフ $\mathcal{S}$ と $\mathcal{S}$ が表す問合せ処理との関係を与える．

[定義3]　スケジュールグラフ $\mathcal{S}=(N, A)$ に基づく問合せ処理とは次の（1）－（3）で定められる手順のことをいう．以下で用いる $R(u)$ （$u \in N$）は直観的にはサイト $\ell(u)$ において得られている関係を表す．

（1）初期設定として，そこに入る有向枝の存在しない $\mathcal{S}$ 上の各節点 $u \in N$ に対し，$\ell(u)=v_i$ ならば $R(u)=r_i$ とする．

（2）各節点 $u \in N$ について，$R(u)$ が求まったならば各有向枝 $(u, w) \in A$ に対しネットワークG上でサイト $\ell(u)$ からサイト $\ell(w)$ へ関係 $R(u)$ を転送する．

（3）各節点 $u \in N$ の $\mathcal{S}$ 上のすべての親節点を w_i（$1 \leq i \leq k$）とする．ネットワークG上でサイト $\ell(u)$ がすべてのサイト $\ell(w_i)$（$1 \leq i \leq k$）から関係 $R(w_i)$ を受取ったならば，サイト $\ell(u)$ は $R(u)=R(w_1) \bowtie R(w_2) \bowtie \cdots \bowtie R(w_k)$ を求める．

単純問合せ $r=r_1 \bowtie r_2 \bowtie \cdots \bowtie r_m$ とスケジュールグラフ $\mathcal{S}=(N, A)$ に対し，$R(u_d)=r_1 \bowtie r_2 \bowtie \cdots \bowtie r_m$ のとき，$\mathcal{S}$ を単純問合せ r に対するスケジュールグラフ

という．

3．応答時間の最小化

ここでは，星状ネットワークGにおける単純問合せに対し，応答時間を最小にする問題RTを考える．

3.1 問題RT

ネットワークGにおける各通信回線の容量C［bit/sec］はすべて等しいとし，X［bit］のデータ転送に要する時間をX／C［sec］と仮定する．

［定義4］ スケジュールグラフ$\mathcal{S}=(\mathcal{N},\mathcal{A})$上の各有向枝$(u,w)\in\mathcal{A}$に対し，重み$t(u,w)$を次の式で与える．

$$t(u,w)=\begin{cases}0 & (\ell(u)=\ell(w))\\ \delta\cdot|R(u)|/C & (\ell(u)\neq\ell(w))\end{cases}$$

ここで，δ［bit］は$R(u)$の1つの組に対応するデータの長さである．

［定義5］ スケジュールグラフ$\mathcal{S}$の応答時間response$(\mathcal{S})$とは$\mathcal{S}$上の重みtに関する最長道の長さである．ここで，道の長さとはその道に含まれる枝の重みの総和と定める．

問題RT

入力：（1）各サイトv_iの関係r_iのサイズ$|r_i|$（$0\leq$

$i \leq n$)

（2） 単純問合せにおける属性Aに対応する値域の要素数D

（3） Dサイト v_d

<u>出力</u>：単純問合せ $r = r_0 \bowtie r_1 \bowtie \cdots \bowtie r_n$ に対するスケジュールグラフ$\mathcal{S}$

<u>目的関数</u>：response（$\mathcal{S}$）→ min

問題RTの解である最適スケジュールにおいては，そのサイズの小さい順に関係を処理すればよいことが示せる(4)．従って，問題RTを解くには関係 r_i（$1 \leq i \leq n$）をそのサイズの小さい順に（$|r_1| \leq |r_2| \leq \cdots \leq |r_n|$とする）並べ換えた後，動的計画法を用いて $r_0 \bowtie r_1 \bowtie \cdots \bowtie r_i$（$1 \leq i \leq n$）に対する最適なスケジュールを，順次，求めて行けばよい．各 r_i に関する計算は $O(n)$ の手数で行えるので，次の定理が得られる．

<u>［定理1］</u> 問題RTは $O(n^2)$ の手数で解ける．但し，nは星状ネットワークにおけるセンタ以外のサイトの数である．

更に，上述のアルゴリズムが出力するスケジュールグラフ$\mathcal{S}$に対し，$\mathcal{S}$は問題RTに対する最適解の中で，特に，通信量の総和が最小のものになっていることが示せる(4)．

4. 通信コスト及び計算コストの総和の最小化

ここでは，星状ネットワークGにおける単純問合せに対し，通信コスト及び計算コストの総和を最小にする問題TCを考える．

4.1 問題TC

ネットワークGの各通信回線において，X [bit] のデータ転送に要する通信コストは $c \cdot X$ であるとする．ここで，c は非負整数とする．次に，各サイトにおいて，関係 r_i と r_j の自然結合 $r_i \bowtie r_j$ の計算に要する計算コストは $p \cdot |r_i| \cdot |r_j|$ であるとする．ここで，p は非負整数とする．

[定義6] スケジュールグラフ $\mathcal{S}=(\mathcal{N}, \mathcal{A})$ 上の各有向枝 $(u, w) \in \mathcal{A}$ に対して $\alpha(u, w)=|R(u)|$ と定め，各節点 $u \in \mathcal{N}$ に対して

$$\beta(u)=\begin{cases} \prod_{(w,u)\in\mathcal{A}} |R(w)| & \left(\begin{array}{l} u\text{に2つ以上の有} \\ \text{向枝が入る場合} \end{array}\right) \\ 0 & (\text{その他}) \end{cases}$$

と定める．

[定義7] スケジュールグラフ $\mathcal{S}=(\mathcal{N}, \mathcal{A})$ の処理コストを

$$\mathrm{total}(\mathcal{S}) = c \cdot \sum_{(u,w)\in\mathcal{A}} \alpha(u, w) + p \cdot \sum_{u\in\mathcal{N}} \beta(u)$$

と定める．

問題TC

入力：（1）各サイト v_i の関係 r_i のサイズ $|r_i|$（$0 \leq i \leq n$）

（2）単純問合せにおける属性Aに対応する値域の要素数D

（3）Dサイト v_d

（4）c，p

出力：単純問合せ $r = r_0 \bowtie r_1 \bowtie \cdots \bowtie r_n$ に対するスケジュールグラフ $\mathcal{S}$

目的関数：$total(\mathcal{S}) \rightarrow min$

問題TCにおいて，$|r_1| \leq |r_2| \leq \cdots \leq |r_n|$ とすると，自然結合の最適な計算順序が $((\cdots((r_0 \bowtie r_1) \bowtie r_2)\cdots) \bowtie r_n)$ となることは容易に示せる(4)．従って，ネスティングの深い方から順に，各自然結合に対する処理のスケジュールを求めればよい．各自然結合に関する計算は定数時間で行えるので，次の定理が得られる．

［定理2］　問題TCは $O(n \log n)$ の手数で解ける．但し，n は星状ネットワークにおけるセンタ以外のサイト数である．

5. あとがき

本稿では，星状ネットワークにおける単純問合せ処理について議論し，応答時間を最小とする処理スケジュールが$O(n^2)$で，通信コスト及び計算コストの総和を最小とする処理スケジュールが$O(n \log n)$でそれぞれ求まることを示した．

一方，通信コストを削減する代表的な手法の1つに，いわゆるデータ圧縮(3)の手法がある．問題TCにおけるサイト間の通信にデータ圧縮の手法を導入すると，自然結合で求まる関係をサイトv_0に転送するのに要する通信コストは無視できるため，問題TCは"関係r_i（$1 \leq i \leq n$）をそのサイズの小さい順に並べ換える問題"と本質的には等価になる．従って，この場合やはり$O(n \log n)$の手数で解ける．

最後に，日頃御指導いただく本学吉田典可教授に深く感謝いたします．

文献

(1) Cheung, T.-Y.:"A method for equijoin queries in distributed relational databases," IEEE Trans. Comput., C-31, 8, pp.746-751(1982).

(2) Hevner, A.R. and Yao, S.B.:"Query processing in distributed database systems," IEEE Trans. Software Eng., SE-5, 3, pp.177-187(1979).

(3) Kambayashi, Y.:"Compressed semi-joins and their applications to distributed query processing," 信学技報, AL81-54, pp.55-62(1981).

(4) 宮尾,杉原:"星状ネットワークにおける単純問合せ処理," ECS Lab., Hiroshima Univ. Tech. Rep. No.83-01(1983).

(5) 杉原,菊野,吉田,宮尾:"分散形データベースシステムにおける単純問合せ処理," 信学技報, AL82-59, pp.21-30(1982).

(6) Ullman, J.D.:"Principles of Database Systems," 2nd ed., Computer Science Press(1982).

(7) Yu, C.T., Chang, C.C. and Chang, Y.C.:"Two surprising results in processing simple queries in distributed databases," Proc. COMPSAC 82, pp.377-384(1982).